

- 78

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

STRUCTURE OF REPEAT BUYING
FOR NEW PACKAGED GOODS*

Manohar U. Kalwani and Alvin J. Silk**

WP 1032-78

October 1978

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

STRUCTURE OF REPEAT BUYING
FOR NEW PACKAGED GOODS*

Manohar U. Kalwani and Alvin J. Silk**

WP 1032-78

October 1978

*The authors are indebted to Jiri Nechleba for his valuable assistance in analysis and interpretation of data. Special thanks are also due to John McFadden for his help in processing the Chicago Tribune Panel tapes and carrying out statistical analyses.

**Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT

A prime concern of marketing managers in making decisions relating to the development and introduction of new packaged goods is the new product's ability to sustain a satisfactory level of repeat buying. This paper reports empirical findings from tests of hypotheses on the structure of repeat buying for new packaged goods set forth by Eskin and Malec. It extends their work by examining the theoretical justification for these hypotheses. A linkage is established between the discrete model underlying the hypotheses and a frequently used continuous time model based on gamma mixture of exponential interpurchase times. These postulates are tested for four new packaged goods using a maximum likelihood approach. Overall, the results obtained support the hypotheses and imply that for a new product, the penetration for various classes of repeat and hence total repeat sales, can be obtained from a knowledge of the cumulative proportion of first repeaters and their average interpurchase time.

INTRODUCTION

A prime concern of marketing managers in making decisions relating to the development and launching of new packaged goods is the new product's ability to sustain a satisfactory level of repeat buying. Diagnostic information on this important issue is routinely obtained from tracking studies of new product test markets and introductions wherein over time purchase data are decomposed into levels of trial and depth of repeat purchasing. A number of forecasting models are available which utilize such measurements as inputs [9, 11, 14, 16]. Does repeat buying of new packaged goods exhibit common structural features which hold for different brands and product categories? Knowledge about such adoption phenomena constitutes the foundation for analytical efforts directed toward the assessment of new brands prior to test marketing [18] and the subsequent forecasting of sales from early test market results. Building upon the ideas put forth in Eskin's depth of repeat model [9], PANPRO, Eskin and Malec [10] have recently reported progress in understanding the process of how repeat buying develops.

The present paper reports some further analysis and empirical findings concerning the structure of repeat buying for new packaged goods. The theoretical rationale for a set of hypotheses on the structure of repeat buying suggested by the work of Eskin [9] and Eskin and Malec [10] is examined. A linkage is established between the underlying discrete depth of repeat model and a well-known continuous time model based on a gamma

mixture of exponential interpurchase times. The depth of repeat model is then used to analyze the structure of repeat buying -- i.e., the conversion of triers into the first repeat class, the conversion of first repeaters into the second repeat class, and so on for higher repeat levels. Our interest is to compare the time paths of these cumulative conversion proportions across various repeat levels. Do these penetration curves have similar functional forms? Are the interpurchase times approximately the same across repeat levels? Does the cumulative proportion of consumers who convert from one repeat class to the next increase with depth of repeat? Model parameters estimated by a maximum likelihood method are used to test hypotheses related to these questions.

The remainder of this paper is organized as follows: The first section contains a description and specification of the depth of repeat model (with time treated discretely) along with a set of hypotheses on the structure of repeat buying suggested by the work of Eskin [9] and Eskin and Malec [10]. Next, we introduce a continuous analog of the discrete model to establish the connection with a type of stochastic model frequently employed by marketing researchers with reference to established products. The presentation of the continuous model is followed by the development of the likelihood expressions to estimate model parameters. Then we present some empirical results for four new products - two toothpastes and two coffees. Finally, we summarize our results and discuss their implications for consumer research concerned with the evaluation of new packaged goods.

DEPTH OF REPEAT MODELS

Discrete Models:

Our analysis of repeat buying follows the work of Fournier and Woodlock [11], Massy [14] and Eskin [9]. Depth of repeat classes are defined as the penetration or cumulative proportion of consumers who ever repeat a J^{th} time ($J=1,2,3,\dots$) given that they had previously made $J-1$ repeat purchases. Note that the repeat class (or level) $J=1$ refers to first repeaters, the repeat class $J=2$ refers to second repeaters, and so on for higher repeat levels. Total repeat purchases are obtained by summing across repeat classes.

Two postulates underly the penetration model proposed by Fournier and Woodlock [12]. First, a ceiling is presumed to exist on the proportion of consumers who ever convert from one repeat class to the next. Secondly, the number entering the next repeat level in each time period is taken to be a constant fraction of the proportion who will yet convert into the next repeat class. Fournier and Woodlock [11, p. 32] cite empirical evidence to support both these postulates:

Observation of numerous annual cumulative penetration curves shows that (1) successive increments to these curves decline, and that (2) the cumulative curves seem to approach a limiting penetration of less than 100 per cent of households -- frequently far less.

Consider repeat level J where $J=1,2,3,\dots$. Given the above two postulates it follows that the proportion of consumers, $P(L)$, who convert into the modeled repeat level J during the L^{th} time period is given by¹:

$$P(L) = F(L) - F(L-1) = (1-p) (\alpha - F(L-1)), \quad L=1,2,3,\dots, \quad (1)$$

where: $F(L)$ represents the cumulative proportion of consumers who convert into a particular repeat level, J , by the L^{th} time period since the previous purchase; ρ is a constant; and α denotes the cumulative proportion who will eventually convert into the repeat class J . Thus, α represents the ceiling on penetration. Solving the above equation recursively (with $F(0) = (0)$), we obtain:

$$P(L) = (1-\rho) (\alpha - F(L-1)) = \alpha \rho^{L-1} (1-\rho), \quad L=1,2,3,\dots \quad (2)$$

$$F(L) = \sum_{L=1}^L \alpha \rho^{L-1} (1-\rho) = \alpha (1-\rho^L), \quad L=1,2,3,\dots \quad (3)$$

Figure 1 provides a graphic display of the penetration function, $F(L)$, over time. An examination of the above expressions for $P(L)$ and $F(L)$ reveals that the purchasing behavior of consumers who enter repeat class J is given by a geometric distribution. The average interpurchase time, τ ,

 INSERT FIGURE 1 HERE

of these consumers who will eventually convert into some repeat level J is given by²:

$$E[L] = \tau = \frac{\sum_{L=1}^{\infty} L \alpha \rho^{L-1} (1-\rho)}{\sum_{L=1}^{\infty} \alpha \rho^{L-1} (1-\rho)} = \left(\frac{1}{1-\rho} \right) \quad (4)$$

Rearranging (4), we obtain:

$$\rho = \left(\frac{\tau-1}{\tau} \right). \quad (5)$$

Note that the expression for interpurchase time given in equation (4) assumes that data on conversions into the J^{th} repeat level are available over an infinite period of time. In practice, however, the period over which conversion rates have actually been observed is bound to be limited.

Assuming that the duration of observation time for conversion into the J^{th} repeat class is T periods, the expression for average interpurchase

time, τ_T becomes:

$$\tau_T = \frac{\sum_{L=1}^T \alpha \rho^{L-1} (1-\rho)}{\sum_{L=1}^T \alpha \rho^{L-1} (1-\rho)} = \left(\frac{1}{1-\rho} \right) - \left(\frac{T \rho^T}{1-\rho^T} \right) . \quad (6)$$

Hence, the average interpurchase time, τ_T , for truncated data is smaller than the theoretical average interpurchase time, τ .

Thus far it has been assumed that the consumers entering a particular repeat level are homogeneous. Empirical evidence, however, indicates that the early entrants into a repeat class tend to be heavier buyers of the product than are later entrants [9, 11, 14, 16]. Fourt and Woodlock [11, pp. 33-34] note that the estimated penetration levels based on equation (3) fit the data well except that the predictions for distant time periods tend to be too low. The poor fit is attributed to heavier buyers converting into a repeat class earlier than light buyers. The addition of a trend factor, δ , provides for this "stretch out" of penetration. This leads to the following adjustments in equations (2) and (3):

$$P(L) = \alpha \rho^{L-1} (1-\rho) + \delta, \quad L=1,2,3,\dots \quad (7)$$

$$F(L) = \alpha (1-\rho^L) + \delta L, \quad L=1,2,3,\dots \quad (8)$$

The preceding modification allows the ceiling to be a linear function of time instead of a fixed quantity. The expression for average interpurchase time given a finite observation time period T now becomes:

$$\tau_T^\delta = \frac{\sum_{L=1}^T \alpha \rho^{L-1} (1-\rho) + \delta}{\sum_{L=1}^T \alpha \rho^{L-1} (1-\rho) + \delta} = \frac{\alpha \left\{ \frac{1-\rho^T}{1-\rho} - T \rho^T \right\} + \frac{\delta T (T+1)}{2}}{\alpha (1-\rho^T) + \delta T} \quad (9)$$

Empirical evidence presented by Fourn and Woodlock [11] and Eskin [9] suggests that the numerical value of the "stretch out" factor, δ , is quite small compared to the conversion proportion term, α . Therefore, the effect of the δ term in equation (9) above is likely to be minor, especially when the observation period over which the purchase data are available is of limited duration (say, 12 or 24 weeks).

Eskin and Malec [10] give the following relationship between the interpurchase time and the parameter, ρ :

$$\tau_T^* = \left(\frac{\rho}{1-\rho} \right). \quad (10)$$

Note that this value of interpurchase time is smaller than the theoretical value given in equation (4). Table 1 displays theoretical and truncated interpurchase times for a variety of true values of the parameter, ρ . The truncated τ_T 's in Table 1 were calculated using equation (6). As expected, the truncated interpurchase times, τ_T 's, increase and approach the theoretical, τ , as the duration of observation time increases.

 INSERT TABLE 1 HERE

Hypothesis on Structure of Repeat Buying

For the discrete time model set forth above, the cumulative proportion of consumers who convert into the j^{th} repeat level within L periods of previous purchase is given by:

$$F_J(L) = \alpha_J(1-\rho_J^L) + \delta_J L, \quad L=1,2,\dots, \quad (11)$$

where the suffix, J , denotes the repeat class and L , the time variable, is measured in terms of time periods since last purchase. As indicated earlier, our interest is to compare the penetration curves represented by equation (11) across various repeat levels. How do the parameters, α , ρ , and δ vary with J ? Eskin [9] proposed three hypotheses regarding the patterns in α , ρ , and δ across repeat levels and tested them with purchase data for six established products.

Eskin's first hypothesis was that for a new product the parameter, ρ_J has approximately the same value across all the repeat levels. This implies that the average interpurchase times are approximately equal across all repeat classes. The average interpurchase time for a particular new product is bound to be larger than the average interpurchase time for the product class which, of course, depends on the size of the offering. Also, the average interpurchase time depends on the proportion of consumers who are completely loyal to the test brand and buy it every time following consumption of the product. The larger the proportion of buyers that are committed to the new product and buy it every time, the smaller will be the average interpurchase time. In testing this hypothesis with data for six established products, Eskin [9, p. 127] found that ρ_J "does not fluctuate excessively nor does it exhibit a strong trend".

A second hypothesis put forth by Eskin [9] states that δ_J takes on the same value across all repeat levels. In this case, Eskin found that δ_J 's "vary in a relatively small range but do tend to exhibit a negative trend over J ." As indicated earlier, both Eskin [9] and Fourt and Woodlock [12] found that the magnitude of the δ term is small across all repeat levels. In other words, in the expression for penetration at the J^{th} repeat class

given by equation (11), the first term -- $\alpha_J(1-\rho_J^L)$ -- forms the major contribution, especially at higher repeat levels.

A third hypothesis suggested by Eskin [9] deals with the conversion proportion terms, α_J 's. Eskin [9] postulated that α_J 's could be obtained through a relationship which follows the geometric distribution according to:

$$\alpha_J = \alpha_{\infty}(1-\gamma^J), \quad J=2,3,\dots, \quad (12)$$

where the limit, α_{∞} , is usually 1 or slightly less than 1. He found that the relationship for conversion proportions, α_J 's, given by equation (12), provided a good fit to the data with the estimates of α_{∞} being less than unity for each of the six established products studied. Fitting equation (12) to a data base consisting of observations for an unspecified cross-section of new products, Eskin and Malec [10] obtained an estimate of 0.636 for the parameter γ . With $\alpha_{\infty}=1$, this implies: $\alpha_2 = 60\%$, $\alpha_3 = 74\%$, $\alpha_4 = 84\%$, $\alpha_5 = 90\%$, etc. What interpretation can be given to this systematic pattern in conversion proportion terms across a variety of new packaged goods?

Consumers try a new product on the basis of its expected performance. A first or trial purchase of a new product is ordinarily followed by numerous usage experiences which enables a consumer to evaluate the product's qualities. A repurchase of the new product, which generally is to the exclusion -- partial or whole -- of other products that the consumer was previously buying, is made only if the consumer is satisfied with the test product relative to the previously-used products. Commitment to the new product increases as it is repurchased again and again, and fewer and fewer customers reject the product at higher repeat levels.

The systematic increase in conversion proportion terms, α_j 's, may be interpreted as a learning phenomenon similar to that discussed by Kuehn [13] with reference to established brands. In Kuehn's model, each consecutive purchase of a brand increases the probability that the next purchase will be the same brand. Here, the purchases of the test brand are not necessarily consecutive. Intervening purchases of other brands will tend to prolong the time until a consumer enters the next repeat class.

What are the practical implications of these three hypotheses? The first hypothesis implies that, for a given new product the mean inter-purchase times (or equivalently, the parameter, ρ_j) of second, third, fourth, etc., repeaters is the same as those of first repeaters. The second hypothesis states that for a new product the parameter, δ_j is constant across repeat levels. Note that the final hypothesis on the systematic pattern in the values of $\alpha_2, \alpha_3 \dots$ is applicable across new packaged goods. Assuming that the magnitude of δ_j is small across repeat levels, the above hypotheses taken together imply that for a new product the penetration for various classes of repeat (see equation (1)) can be obtained from a knowledge of the average interpurchase times of first repeaters (or equivalently, ρ_1) and the cumulative proportion of first repeaters, α_1 . Further, to the extent that the interpurchase time of a new brand resembles those of existing brands in the product class, the cumulative repeat proportions for various repeat classes can be estimated solely from knowledge of the cumulative proportion of first repeaters.

Continuous Case:

Consideration of the continuous analog of the discrete depth of repeat model discussed above serves to establish a linkage with a familiar model often applied to purchase data for established brands, namely Ehrenberg's NBD model [5] which assumes exponentially distributed interpurchase times (or equivalently, a Poisson distribution of purchase events across successive time periods of equal length) for an individual household and gamma heterogeneity over the population. Subsequently, we will be interested in comparing the fit of the discrete and continuous models and the marketing interpretation that can be given to their parameter estimates.

As in the discrete case, we assume first that the consumers are homogeneous and that there is a ceiling to the number of consumers who convert from one repeat level to the next. Consider repeat class J and let M denote the proportion of consumers who eventually convert from the repeat level $J-1$ to J where $J=1,2,3,\dots$. The probability of a consumer converting into repeat level J in the time interval, t to $t + h$, given that this consumer has not converted into the J^{th} repeat class in time, t , since last purchase, is taken as:

$$\frac{F(t + h) - F(t)}{M - F(t)} = \lambda h, \quad (13)$$

where M is the ceiling factor, $F(t)$ is the cumulative probability of converting into J^{th} repeat class by time, t , and λ represents the "instantaneous" purchase rate. This postulate in the continuous model is similar to the postulate in the discrete model given in equation (1). Taking the limit as $h \rightarrow 0$ in equation (13) above, we find:

$$\frac{dF(t)}{dt} = f(t) = \lambda(M - F(t)) .$$

Solving this equation by separation of variables, we obtain:

$$F(t) = M(1 - e^{-\lambda t}), \quad (14)$$

and

$$f(t) = M\lambda e^{-\lambda t}, \quad (15)$$

where $f(t)$ is the density function for time to conversion into the j^{th} repeat class. The above two equations reveal that following our two postulates represented in mathematical form in equation (13), the waiting time (or, the interpurchase time) distribution is negative exponential.

The average interpurchase time is given by:

$$\tau = E[t] = \frac{\int_0^{\infty} t M \lambda e^{-\lambda t} dt}{\int_0^{\infty} M \lambda e^{-\lambda t} dt} = \frac{1}{\lambda}. \quad (16)$$

Comparing the expressions for cumulative penetrations in the discrete and continuous cases given in equations (3) and (14), we find:

$$\rho = e^{-\lambda} = e^{-(\frac{1}{\tau})},$$

that is,

$$\tau = \left(\frac{1}{\ln 1/\rho} \right). \quad (17)$$

Anscombe [2] has pointed out that the provision of a ceiling factor, M , is equivalent to assuming that a proportion, M , of the population have a non-zero purchase rate, λ , and a proportion $(1-M)$ have a zero purchase rate. It is reasonable to expect that the fit to empirical data can be improved by assuming a continuous form of heterogeneity. Anscombe [2] and Ehrenberg [5] both employed the Pearson type III or gamma distribution for this purpose which has been shown to work well for consumer purchase data

(see [8]). The expressions for the density function of the gamma distribution along with its mean and variance are given as follows:

$$g(\lambda; \mu, \nu) = \frac{\mu}{\Gamma(\nu)} e^{-\lambda\mu} (\lambda\mu)^{\nu-1}, \quad \lambda > 0,$$

$$E[\lambda] = \frac{\nu}{\mu},$$

and

$$\text{VAR}[\lambda] = \frac{\nu}{\mu^2}.$$

Under the assumptions that the individual waiting time distribution is negative exponential with parameter, λ , and the mixing distribution of λ is gamma, the expressions for $f(t)$ and $F(t)$ are found to be³:

$$f(t) = \frac{\nu}{\mu} \left(\frac{\mu}{\mu+t} \right)^{\nu+1}, \quad (18)$$

and,

$$F(t) = 1 - \left(\frac{\mu}{\mu+t} \right)^{\nu}. \quad (19)$$

Given the waiting time distributions in equations (18) and (19), it is easy to show that the number of purchases, $P_{t^*}(n)$, in the fixed time interval, t^* , is distributed negative binomial:

$$P_{t^*}(n) = \frac{\Gamma(n+\nu)}{n! \Gamma(\nu)} \left(\frac{\mu}{\mu+t^*} \right)^{\nu} \left(\frac{t^*}{\mu+t^*} \right)^n, \quad n=0,1,2,\dots \quad (20)$$

with,

$$E[n] = \frac{\nu t^*}{\mu},$$

and

$$\text{VAR}[n] = \frac{\nu t^* (\mu+t^*)}{\mu^2}.$$

Note, however, that the frequency distribution of the number of purchases obtained in the above manner assumes a stable set of values for μ and ν during the fixed time interval, t^* . Recall that following a depth of repeat approach, there will be a separate values of μ and ν for each repeat class.

A final refinement of the continuous model is to allow for a "zero group" who are not in the market for the new product and are excluded from the "relevant population." Chatfield, Ehrenberg, and Goodhardt [4] found that the negative binomial distribution fails to fit the tails of the observed frequency distribution for number of purchases in a fixed time interval for many populations. Massy, Montgomery, and Morrison [15, pp. 337-338] suggest that this difficulty stems from the fact that the gamma is a unimodal distribution. In cases where there are large numbers of consumers with zero purchase rates, the fitted gamma distribution will tend to take an exponential form. This prevents the distribution from having another mode at $\lambda > 0$, as shown by the dotted curve in Figure 2 which is reproduced here from [15, p. 338].

 INSERT FIGURE 2 HERE

Where the "zero group" is large, it dominates the fitted gamma distribution, which will go to zero rapidly as λ increases. This results in an under-estimation of the proportion of the population with large λ , and can account for the "variance discrepancy" noted by Chatfield, Ehrenberg, and Goodhardt [4]. Massy et. al. [15, p. 237] argue that the gamma heterogeneity should apply only to the relevant population. Thus they suggest that a proportion, A , of the population be modeled as having non-zero purchase rates and that these rates be distributed gamma over the relevant population as shown in Figure 2. The expressions for $f(t)$ and $F(t)$ with this form of heterogeneity are given by

$$f(t) = A \frac{v}{\mu} \left(\frac{\mu}{\mu + t} \right)^{v+1}, \quad (21)$$

and,

$$F(t) = A\{1 - (\frac{\mu}{\mu+t})^v\}. \quad (22)$$

Below we report maximum likelihood estimates of the constrained (μ and v) and the unconstrained (A, μ and v) models and results from likelihood ratio tests for the presence of the aforementioned "spike."

PARAMETER ESTIMATION

Maximum likelihood methods were used to estimate model parameters for each repeat level. It is well-known (see Rao [17]) that under quite general regularity conditions, maximum likelihood estimates are best asymptotically normal (BAN). That is, they are consistent, asymptotically normal, and asymptotically efficient. In addition to these properties, maximum likelihood estimates are invariant. Eskin [9] employs a least squares approach to estimate model parameters and test the three hypotheses stated earlier. The fact that the dependent variable in the regression equation is defined as a cumulative proportion, and hence is nondecreasing, makes the presumption of uncorrelated error terms tenuous. In the presence of autocorrelation, Eskin [9] used a generalized least square procedure to obtain consistent parameter estimates. In some preliminary analyses of the products studied here, the same generalized least squares approach followed by Eskin was used to estimate parameters for the discrete depth of repeat model. However, results obtained were unsatisfactory. Estimates of α that exceeded unity were found for some repeat levels, and some estimates of δ were negative. Measurement errors in the penetration data might account for these difficulties. The estimates were also unstable due to small sample sizes for two of the new products investigated, especially at higher repeat levels.

The difficulty with employing the maximum likelihood method is that it is not possible to obtain closed form analytical solutions. Therefore numerical optimization is required to obtain the maximum likelihood estimates. For this purpose, use was made of a general optimization procedure developed by Kalwani [12]. The computer program (written in FORTRAN IV) executes an accelerated pattern search technique. Note that the optimization procedure permits the search to be restricted to the feasible range of solutions. However, we did not find it necessary to impose such restrictions for any of the four new products.

Discrete Model

The likelihood expression for the discrete model, $\ell(D)$, applicable to any one of the repeat levels (say, J), is given by:

$$\ell(D) = \prod_{L=1}^{52} (1-F(L))^{\bar{n}_L} \prod_{L=1}^{52} (P(L))^{n_L}, \quad (23)$$

where:

$$F(L) = \alpha(1-\rho^L) + \delta L,$$

$$P(L) = \alpha\rho^L(1-\rho) + \delta,$$

and,

n_L = Number of consumers who enter repeat level J within L time units since their previous purchase,

\bar{n}_L = Number of consumers who had L time units available to convert into the J^{th} repeat level but did not do so.

Note that $\sum_{L=1}^{52} n_L = n$ represents the total number of consumers who converted into a given repeat level within 52 weeks of a previous purchase. Similarly, $\sum_{L=1}^{52} \bar{n}_L = (m-n)$ denotes the total number of consumers who have not entered

the current repeat level given that altogether, m consumers made $(J-1)$ purchases of the new product.

Continuous Model

Consider first the constrained version of the continuous model. Recall from the previous section that in this case all buyers at the previous repeat level (say, $J-1$) are included in the "relevant population" and modeled for possible entry into the level J . In mathematical terms, we have constrained the parameter A to be zero. The likelihood expression, $\ell(C)$, for repeat level J where $J=1,2,\dots$ is given below.

$$\ell(C) = \prod_{L=1}^{52} (1-F(L))^{\bar{n}_L} \prod_{L=1}^{52} (F_L - F_{L-1})^{n_L}, \quad (24)$$

where,

$$F(L) = 1 - \left(\frac{\mu}{\mu+L}\right)^\nu,$$

and \bar{n}_L , and n_L are defined in the same way as for the discrete model.

In the case of the unconstrained version of the continuous model, the only change is in the expression for $F(L)$ which is altered to:

$$F(L) = A\{1 - \left(\frac{\mu}{\mu+L}\right)^\nu\}.$$

A likelihood ratio test may be employed to determine which of the two models -- constrained or unconstrained -- provides a statistically superior fit to the data. We form a ratio of the maximum value of the likelihood function of the constrained model (null hypothesis) to that for the unconstrained model (alternate hypothesis):

$$\frac{\ell^*(C)}{\ell^*(UC)} = R,$$

where $\ell^*(C)$ denotes the maximum value of the likelihood function for the constrained model and $\ell^*(UC)$ denotes the maximum value of the likelihood

function for the unconstrained model. For large samples, Wilks [19] has shown that $-2\log_e R$ is distributed chi-square (with 1 degree of freedom in this case).

FINDINGS

Data Base

The results reported below are based on an analysis of purchase data for four new products, two brands of toothpaste -- UltraBrite and Plus White -- and two brands of freeze-dry coffees -- Maxim and Taster's Choice. The source of purchase information is the panel data collected by Chicago Tribune's Family Survey Bureau. This panel contains purchase records for 530 households. The largest amount of purchase information was available in the case of UltraBrite where the observations covered a three-year period following its introduction. In the cases of Plus White and the two brands of coffees, the purchase data extended over a period of about 2 years following the introduction of each of these products. We have reported findings here for only those repeat levels where the sample sizes are at least 30. Maximum likelihood estimates of parameters are displayed in Tables 2 - 5 for the discrete models, and Tables 6 - 9 for the continuous version of the model.

INSERT TABLES 2 - 5 HERE

Discrete Model:

Consider first the parameter estimates for the discrete case and the three hypotheses developed by Eskin and discussed earlier concerning repeat buying patterns. The first hypothesis was that the parameter ρ

would be approximately the same across repeat levels for each of the four products. In the case of Ultra Brite, where the sample size is largest, the results displayed in Table 2 reveal very little variation across the four repeat levels with the estimated ρ 's fluctuating around .94. The results for Plus White (Table 3) and the two coffees (Table 4 and 5) appear less consistent; but the variability is still slight, the ranges of the estimated ρ 's for these three products being as follows: Plus White, .89-.93; Maxim, .84-.92; and Taster's Choice, .88-.92. It should be pointed out that in the case of the two coffees we would expect some variation in ρ which is related to the interpurchase time, due to seasonal variations in the consumption of coffee.

A second hypothesis was that the parameter δ would be approximately the same across repeat levels for each of the four products. Our results show that the estimated values of δ are small in magnitude, generally less than 0.2 per cent. This finding is consistent with the empirical experience of Fourt and Woodlock [11].

The final hypothesis is to be considered relates to the conversion proportion terms, α_j 's. As noted previously, on the basis of the empirical results Eskin and Malec reported for a cross-section of new products, it was postulated that these terms systematically increase to α_∞ which is usually slightly less than .1. More specifically, given Eskin and Malec's estimate of $\hat{\gamma} = .636$ for (12) it was expected that the α_j 's would take on the following values: $\alpha_2 = 60\%$, $\alpha_3 = 74\%$, $\alpha_4 = 84\%$, etc. The maximum likelihood estimates of the α_j terms for the four new products are displayed in Tables 2 through 5. Examining the estimates of α_2 , α_3 , and α_4 for Ultra Brite displayed in Table 2, we find that while the estimates of α_j 's are

not exactly equal to the hypothesized values, they do exhibit a non-decreasing pattern. Table 6 displays the differences between estimated and hypothesized values of the α_j 's. Note the deviations never exceed approximately 10 per cent of their hypothesized values.

INSERT TABLE 6 HERE

Continuous Model

The results obtained for the continuous depth of repeat model are presented in Tables 7 through 10. On examination of the test statistic, $-2\ln R$, for the likelihood ratio test we conclude that the null hypothesis that the constrained model provides a better fit can be rejected at the .10 level for all four products. Earlier we indicated that the poor fit of the two parameter gamma (i.e., constrained model) could be attributed to the possible presence of two modes in the distribution of purchase rates over the population -- one at zero and the other at non-zero value.

INSERT TABLES 7 - 10 HERE

Finally, we present a comparison of penetration estimates from the better fitting continuous model (i.e., unconstrained version) with those from the discrete model. We have chosen UltraBrite to illustrate our results since this is the case for which the largest amount of purchase information was available for estimation purposes.

INSERT TABLE 11 HERE

Table 11 displays the penetration estimates at the end of 12, 24, 36, or 52 weeks for each of the first four repeat levels. The estimated penetrations based on the discrete and unconstrained version of the continuous models are very close to the actual penetration levels, particularly for the 52 week period. Note that the results shown in Table 11 relate to the goodness of fit rather than the predictive accuracy of the discrete and continuous models. Our conclusion from these findings is that the discrete model with parameters that have marketing interpretation provides a fit as good as the unconstrained continuous model. Recall that the two key parameters of the discrete model are ρ and α . As indicated earlier, the parameter, ρ is directly related to average interpurchase time, and the parameter, α simply represents the cumulative proportion of consumers who convert from a given repeat level to the next.

DISCUSSION AND CONCLUSIONS

In this paper, we have reported results from some empirical tests of three hypotheses set forth by Eskin [9] and Eskin and Malec [10] on the structure of repeat buying. Maximum likelihood estimates of model parameters were developed for four new packaged goods - two toothpastes and two coffees and compared for consistency with the hypothesized patterns. The first hypothesis predicted that for a given new product the parameter ρ_j , a measure of average interpurchase time, would be constant across repeat levels. Overall, the parameter estimates obtained were found to be consistent with this hypothesis -- especially in the case

of Ultra Brite and to a lesser but still supportive degree for the other three new brands. As hypothesized, the parameter δ_j displayed little variation across repeat levels and further, its magnitude was generally found to be very small.

The empirical tests of the final hypothesis on systematic patterns in cumulative conversion proportion terms $\alpha_2, \alpha_3 \dots$ revealed that while the estimates of these parameters deviated somewhat from their hypothesized values, they did exhibit the postulated non-decreasing pattern (see Table 6 for differences between estimates and hypothesized values of α_j 's).

While these results are generally encouraging, attention needs to be drawn to the limited number of repeat levels that were used here to test these hypotheses. In spite of the fact that the purchase records for each of the four products extended over a period of at least two years, the small size of the Chicago Tribune Panel yielded only a few repeat levels with sample sizes of 30 or more.

The aforementioned hypotheses carry important implications for the tasks of making either pre-test market or early test market forecasts of the time path and equilibrium level of penetration for new brands. Given that these hypotheses about repeat buying hold, it follows that penetration levels for various repeat classes depend primarily on two factors: the average interpurchase time of first repeaters (or equivalently the parameter ρ_1) and the cumulative proportion of first repeaters. A reasonable initial estimate of ρ_1 (or τ_1) could be obtained for a new brand by examining the interpurchase times of existing brands in the product class. In devising such an account would obviously need to be taken of any differences in the new and existing brands' package sizes. Then, the repeat sales for the new brand could be

forecast from a knowledge of the cumulative proportion of triers who repeat buy it at least once. Research is underway to link these hypotheses about repeat buying to measurement methodologies used to assess new brands prior to test marketing [18].

FOOTNOTE

1. Note that in equation (1) the number of consumers purchasing the new packaged good (or the number of buyers entering a particular repeat class) is not influenced by the number who have already purchased the new product (i.e., $F(L-1)$).
2. The expression for variance can be obtained in a manner similar to the derivation of expected value of L .

$$\text{VAR}[L] = \frac{\sum_{L=1}^{\infty} L^2 \alpha \rho^{L-1} (1-\rho)}{\sum_{L=1}^{\infty} \alpha \rho^{L-1} (1-\rho)} - (E[L])^2 = \frac{\rho}{(1-\rho)^2} .$$

Table 1 displays the expected values of L (i.e., theoretical τ) for three illustrative values of ρ . For $\rho = 5/6$, $7/8$, and $9/10$ the expectation of L equals 6, 8, and 10, respectively and the corresponding values of variance of L are 30, 56, and 90.

3. The expression for average interpurchase time in the continuous case is given by:

$$E[t] = \int_0^{\infty} t f(t) dt = \int_0^{\infty} t \frac{v}{\mu} \left(\frac{\mu}{\mu+t} \right)^{v+1} dt = \left(\frac{\mu}{v-1} \right) .$$

Table 1

THEORETICAL AND OBSERVED INTERPURCHASE TIMES IN WEEKS

TRUE VALUE OF ρ	THEORETICAL τ (Equation (4))	OBSERVED τ_T 's (Equation (6))				τ_T (Equation (10))
		T=12 WEEKS	T=24 WEEKS	T=36 WEEKS	T=52 WEEKS	
5/6	6	4.48	5.69	5.95	6.00	5
7/8	8	4.97	7.17	7.70	7.95	7
9/10	10	5.28	8.20	9.17	9.78	9

Table 2

DISCRETE MODEL: ULTRA BRITE TOOTHPASTE

REPEAT LEVEL	SAMPLE SIZE	PARAMETER ESTIMATES		
		α	ρ	δ
1	134	.5410	.9394	.0001
2	86	.6786	.9474	.0000
3	66	.7116	.9446	.0018
4	49	.7560	.9363	.0000

Table 3

DISCRETE MODEL: PLUS WHITE TOOTHPASTE

REPEAT LEVEL	SAMPLE SIZE	PARAMETER ESTIMATES		
		α	ρ	δ
1	64	.3929	.9264	.0000
2	39	.5378	.8928	.0014
3	31	.7991	.9042	.0000

Table 4

DISCRETE MODEL: MAXIM COFFEE

REPEAT LEVEL	SAMPLE SIZE	PARAMETER ESTIMATES		
		α	ρ	δ
1	118	.3027	.8394	.0025
2	75	.6191	.8837	.0014
3	54	.7605	.9190	.0000

Table 5

DISCRETE MODEL: TASTER'S CHOICE COFFEE

REPEAT LEVEL	SAMPLE SIZE	PARAMETER ESTIMATES		
		α	ρ	δ
1	150	.6425	.9235	.0000
2	99	.6694	.8852	.0014
3	72	.8092	.8755	.0003

Table 6

DIFFERENCES BETWEEN ESTIMATES AND HYPOTHESIZED VALUE OF α_j 'S

PARAMETER	HYPOTHESIZED VALUE	OBSERVED-HYPOTHESIZED VALUE			
		ULTRA- BRITE	PLUS WHITE	MAXIM	TASTER'S CHOICE
α_2	.60	.0786	-.0622	.0191	.0694
α_3	.74	-.0284	-.0409	.0205	.0692
α_4	.84	-.0840	n.a.	n.a.	n.a.

n.a. denotes "not available".

Table 7

CONTINUOUS MODEL: ULTRA BRITE TOOTHPASTE

REPEAT LEVEL	CONSTRAINED		UNCONSTRAINED			$-2 \ln R^a$
	μ	ν	μ	ν	A	
1	12.71	0.47	190.1	11.79	0.560	9.35
2	20.80	0.83	168.2	9.01	0.704	2.55
3	37.61	1.69	87.43	4.23	0.891	0.22
4	20.32	1.08	215.0	14.38	0.765	2.83

$Pr(x_1^2 \geq 2.7) = 0.1$

Table 8

CONTINUOUS MODEL: PLUS WHITE TOOTHPASTE

REPEAT LEVEL	CONSTRAINED		UNCONSTRAINED			$-2 \ln R^a$
	μ	ν	μ	ν	A	
1	5.65	0.21	221.1	17.27	0.396	7.26
2	6.79	0.45	61.11	5.79	0.627	3.70
3	11.84	1.03	269.1	27.68	0.801	4.68

$$^a \Pr(x_1^2 \geq 2.7) = 0.1$$

Table 9

CONTINUOUS MODEL: MAXIM COFFEE

REPEAT LEVEL	CONSTRAINED		UNCONSTRAINED			$-2 \ln R^{\text{a}}$
	μ	ν	μ	ν	A	
1	2.91	0.20	6.79	0.74	0.543	1.93
2	5.46	0.52	50.12	5.49	0.703	5.41
3	6.30	0.63	10.21	1.09	0.872	0.17

$$^{\text{a}}\Pr(x_1^2 \geq 2.7) = 0.1$$

Table 10

CONTINUOUS MODEL: TASTERS CHOICE COFFEE

REPEAT LEVEL	CONSTRAINED		UNCONSTRAINED			$-2 \ln R^{\textcircled{a}}$
	μ	ν	μ	ν	A	
1	8.17	0.53	66.53	5.62	0.659	5.79
2	7.02	0.69	71.21	7.81	0.744	4.46
3	6.31	0.92	23.76	3.51	0.846	1.23

$$\textcircled{a} \Pr(x_1^2 \geq 2.7) = 0.1$$

Table 11

OBSERVED AND ESTIMATED PENETRATIONS (%) FOR ULTRA BRITE

REPEAT LEVEL	T = 12 WEEKS			T = 24 WEEKS			T = 36 WEEKS			T = 52 WEEKS		
	OB- SER- VED	ESTIMATED		OB- SER- VED	ESTIMATED		OB- SER- VED	ESTIMATED		OB- SER- VED	ESTIMATED	
		DIS.	CONT.		DIS.	CONT.		DIS.	CONT.		DIS.	CONT.
1	30.3	28.7	28.8	43.3	42.2	42.2	48.8	48.8	48.8	52.8	52.5	52.8
2	32.8	32.4	32.6	50.7	49.3	49.2	58.2	58.2	58.1	64.2	63.8	64.2
3	36.0	37.4	37.4	59.3	57.4	57.2	67.4	68.5	68.4	76.7	76.8	76.7
4	43.3	41.3	41.5	62.7	60.0	59.8	68.7	68.5	68.2	73.1	73.1	73.1

Figure 1

PENETRATION FOR REPEAT LEVEL J

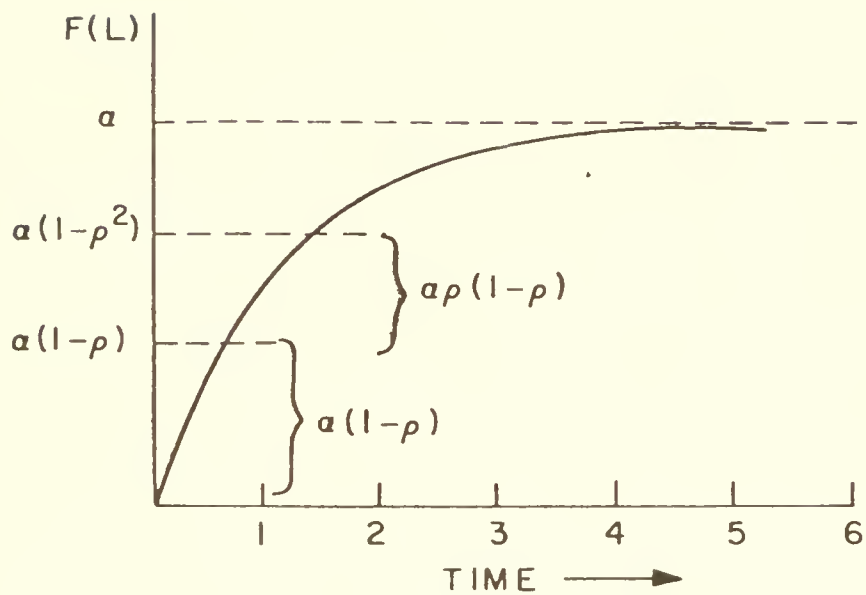
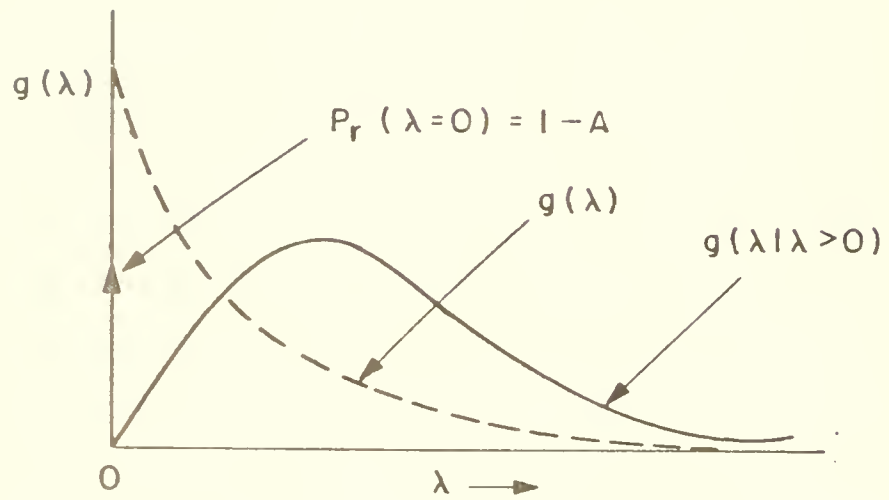


Figure 2

DISTRIBUTION OF PURCHASE RATES



REFERENCES

1. Ahl, David H. "New Product Forecasting Using Consumer Panels," Journal of Marketing Research, 7 (May 1970), 160-7.
2. Anscombe, F.J. "Estimating a Mixed-Exponential Response Law," Journal of the American Statistical Association, 56 (September 1961), 493-502.
3. Baum, J. and K.E.R. Dennis. "The Estimation of the Expected Brand Share of a New Product," ESOMAR Congress, 1961.
4. Chatfield, C., A.S.C. Ehrenberg, and G.J. Goodhardt. "Progress on a Simplified Model of Stationary Purchasing Behavior," Journal of the Royal Statistical Society A, 129 (1966), 317-67.
5. Ehrenberg, A.S.C. "The Pattern of Consumer Purchases," Applied Statistics, 8 (1959), 26-41.
6. Ehrenberg, A.S.C., and G.J. Goodhardt. "Repeat Buying of a New Brand -- A 10-point Case History," British Journal of Marketing, 2 (Autumn 1968), 200-5.
7. Ehrenberg, A.S.C. "Predicting the Performance of New Brands," Journal of Advertising Research, 11 (December 1971), 3-10.
8. Ehrenberg, A.S.C. Repeat Buying. North Holland, Amsterdam, 1972.
9. Eskin, Gerald J. "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," Journal of Marketing Research, 10 (May 1973), 115-29.
10. Eskin, Gerald J., and John Malec. "A Model for Estimating Sales Potential Prior to the Test Market," Proceedings, 1976 Educators' Conference, Series No. 39, American Marketing Association, 220-33.
11. Fourn, Louis A., and Joseph W. Woodlock. "Early Prediction of Market Success for New Grocery Products," Journal of Marketing, 24 (October 1960), 31-8.
12. Kalwani, M.U. "Zero-Order Models: Specification, Data Requirements, and Estimation," Unpublished Ph.D. dissertation, Graduate School of Business, Columbia University, 1975.
13. Kuehn, A.A. "Consumer Brand Choice As A Learning Process" , Journal of Advertising Research, 2, (December 1962), 10-17.
14. Massy, William F. "Forecasting the Demand for New Convenience Products," Journal of Marketing Research, 6 (November 1969), 405-12.

15. Massy, William F., David B. Montgomery, and Donald G. Morrison. Stochastic Models of Buyer Behavior. Cambridge, MA: The M.I.T. Press, 1970.
16. Parfitt, J.H., and B.J.K. Collins. "Use of Consumer Panels for Brand-Share Prediction," Journal of Marketing Research, 5 (May 1968), 131-45.
17. Rao, C.R. Linear Statistical Inference and its Applications. New York: John Wiley & Sons, 1965.
18. Silk, Alvin J. and Glen L. Urban, "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15 (May 1978), 171-191.
19. Wilks, Samuel S. Mathematical Statistics. New York: John Wiley & Sons, 1962.

JY 04 '89

MAR 30 '89

MAR 28 '89

MAR 31 1990

SEP 27 1991

ACME
BOOKBINDING CO., INC.

SEP 6 1983

100 CAMBRIDGE STREET
CHARLESTOWN, MASS.

MT - LIBRARY
3 9080 004 524 481

MT - LIBRARY
3 9080 004 524 507

1031-78

HD28.M414 no.1032- 78
Kalwani, Manoh/Structure of repeat buy
736263 D*BKS 00069315
3 9080 001 240 651

HD28.M414 no.1032- 78 1979
Kalwani, Manoh/Structure of repeat buy
737571 D*BKS 00137169
3 9080 002 050 018

HD28.M414 no.1033- 78
Baldwin, Carl/Liquidity preference un
736011 D*BKS 00066623
3 9080 001 209 383

HD28.M414 no.1034- 78
Van Breda, Mic/Bayes vs. the lens :
736013 D*BKS 00066626
3 9080 001 209 433

HD28.M414 no.1035- 78
Simon, Hermann/Dynamics of price elast
736015 D*BKS 00066627
3 9080 001 209 458

HD28.M414 no.1036- 78
Schmalensee, R/A simple model of risk
736017 D*BKS 00066625
3 9080 001 209 417

HD28.M414 no.1038- 79
Choffray, Jean/Methodology for segment
736019 D*BKS 00066621
3 9080 001 209 342

HD28.M414 no.1039- 79
Meldman, Jeffr/Decision support system
827646 D*BKS 00069325
3 9080 001 240 875

HD28.M414 no.1040- 79
Kaplan, Nancy /A computer code for sol
736021 D*BKS 00066622
3 9080 001 209 367

